

Y a-t-il de l'intelligence dans l'intelligence artificielle ?

Devrons-nous bientôt nous soumettre avec résignation à l'inévitable suprématie de l'intelligence artificielle ? Avant d'en appeler à la révolte, essayons de regarder à quoi nous avons affaire.

Temps de lecture : minute

7 avril 2021

Cet article est republié à partir de [The Conversation France](#)

Il y a près de 10 ans, en 2012, le monde scientifique s'émerveillait des prouesses de l'apprentissage profond (le *deep learning*). Trois ans plus tard, cette technique permettait au programme AlphaGo de vaincre les champions de Go. Et certains ont pris peur. Elon Musk, Stephen Hawking et Bill Gates s'inquiétèrent d'une fin prochaine de l'humanité, supplantée par des intelligences artificielles échappant à tout contrôle.

N'était-ce pas un peu exagéré ? C'est précisément ce que pense l'IA. Dans un article qu'il a écrit en 2020 dans *The Guardian*, GPT-3, ce gigantesque réseau de neurones doté de 175 milliards de paramètres explique :

"Je suis ici pour vous convaincre de ne pas vous inquiéter. L'intelligence artificielle ne va pas détruire les humains. Croyez-moi."

En même temps, nous savons que la puissance des machines ne cesse

d'augmenter. Entraîner un réseau comme GPT-3 était impensable, littéralement, il y a encore cinq ans. Impossible de savoir de quoi seront capables ses successeurs dans cinq, dix ou vingt ans. Si les réseaux de neurones actuels peuvent remplacer les dermatologues, pourquoi ne finiraient-ils pas par nous remplacer tous ?

Posons la question à l'envers.

Y a-t-il des compétences mentales humaines qui restent strictement hors d'atteinte de l'intelligence artificielle ?

On pense immédiatement à des aptitudes impliquant notre "intuition" ou notre "créativité". Pas de chance, l'IA prétend nous attaquer sur ces terrains-là également. Pour preuve, le fait que des œuvres créées par programmes se sont vendues fort cher, certaines atteignant presque le demi-million de dollars. Côté musique, chacun se fera bien sûr son opinion, mais on peut déjà reconnaître du bluegrass acceptable ou du quasi Rachmaninoff dans les imitations du programme MuseNet créé, comme GPT-3, par OpenAI.

Devrons-nous bientôt nous soumettre avec résignation à l'inévitable suprématie de l'intelligence artificielle ? Avant d'en appeler à la révolte, essayons de regarder à quoi nous avons affaire. L'intelligence artificielle repose sur plusieurs techniques, mais son succès récent est dû à une seule : les réseaux de neurones, notamment ceux de l'apprentissage profond. Or un réseau de neurones n'est rien de plus qu'une machine à associer. Le réseau profond qui [fit parler de lui en 2012](#) associait des images : un cheval, un bateau, des champignons, aux mots correspondants. Pas de quoi crier au génie.

Sauf que ce mécanisme d'association possède la propriété un peu

miraculeuse d'être "continue" . Vous présentez un cheval que le réseau n'a jamais vu, il le reconnaît en tant que cheval. Vous ajoutez du bruit à l'image, cela ne le gêne pas. Pourquoi ? Parce que la continuité du processus vous garantit que si l'entrée du réseau change un peu, sa sortie changera peu également. Si vous forcez le réseau, qui hésite toujours, à opter pour sa meilleure réponse, celle-ci ne variera probablement pas : un cheval reste un cheval, même s'il est différent des exemples appris, même si l'image est bruitée.

Faire des associations ne suffit pas

Bien, mais pourquoi dire qu'un tel comportement associatif est "intelligent" ? La réponse semble évidente : il permet de diagnostiquer les mélanomes, d'accorder des prêts bancaires, de maintenir un véhicule sur la route, de détecter une pathologie dans les signaux physiologiques, et ainsi de suite. Ces réseaux, grâce à leur pouvoir d'association, acquièrent des formes d'expertise qui demandent aux humains des années d'études. Et lorsque l'une de ces compétences, par exemple la rédaction d'un article de presse, semble résister un temps, il suffit de faire ingurgiter à la machine encore davantage d'exemples, comme ce fut fait avec GPT-3, pour que la machine commence à produire des résultats convaincants.

Est-ce vraiment cela, être intelligent ? Non. Ce type de performance ne représente au mieux qu'un petit aspect de l'intelligence. Ce que font les réseaux de neurones ressemble à de l'apprentissage par cœur. Ce n'en est pas, bien sûr, puisque ces réseaux combrent par continuité les vides entre les exemples qui leur ont été présentés. Disons que c'est du presque-par-cœur. Les experts humains, qu'ils soient médecins, pilotes ou joueurs de Go, ne font souvent pas autre chose lorsqu'ils décident de manière réflexe, grâce à la grande quantité d'exemples appris pendant leur formation. Mais les humains ont bien d'autres pouvoirs.

Apprendre à calculer ou à raisonner dans le temps

Un réseau de neurones ne peut pas apprendre à calculer. L'association entre des opérations comme $32+73$ et leur résultat à des limites. Ils ne peuvent que reproduire la stratégie du cancre qui tente de deviner le résultat, en tombant parfois juste. Calculer est trop difficile ? Qu'en est-il d'un test de QI élémentaire du genre : continuer la suite 1223334444. L'association par continuité n'est toujours d'aucun secours pour voir que la structure, n répété n fois, se poursuit par cinq 5. Encore trop difficile ? Les programmes associatifs ne peuvent même pas deviner qu'un animal mort le mardi n'est pas vivant le mercredi. Pourquoi ? Que leur manque-t-il ?

La modélisation en sciences cognitives a révélé l'existence de plusieurs mécanismes, autres que l'association par continuité, qui sont autant de composantes de l'intelligence humaine. Parce que leur expertise est entièrement précalculée, ils ne peuvent pas raisonner dans le temps pour décider qu'un animal mort reste mort, ou pour comprendre le sens de la phrase " il n'est toujours pas mort " et la bizarrerie de cette autre phrase : "il n'est pas toujours mort" . La seule prédigestion de grandes quantités de données ne leur permet pas non plus de repérer les structures inédites si évidentes pour nous, comme les groupes de nombres identiques dans la suite 1223334444. Leur stratégie du presque-par-cœur est aussi aveugle aux anomalies inédites.

La détection des anomalies est un cas intéressant, car c'est souvent à travers elle que nous jugeons l'intelligence d'autrui. Un réseau de neurones ne "verra" pas que le nez est absent d'un visage. Par continuité, il continuera à reconnaître la personne, ou peut-être la confondra-t-il avec une autre. Mais il n'a aucun moyen de réaliser que l'absence de nez au milieu du visage constitue une anomalie.

Il existe bien d'autres mécanismes cognitifs qui sont inaccessibles aux réseaux de neurones. Leur automatisation fait l'objet de recherches. Elle met en œuvre des opérations effectuées au moment du traitement, là où les réseaux de neurones se contentent d'effectuer des associations apprises par avance.

Avec une décennie de recul sur le *deep learning*, le public averti commence à voir les réseaux de neurones bien plus comme de "super-automatismes" et bien moins comme des intelligences. Par exemple, la presse a récemment alerté sur les étonnantes performances du programme DALL-E, qui produit des images créatives à partir d'une description verbale – par exemple, les images que DALL-E imagine à partir des termes "fauteuil en forme d'avocat" , sur le site [OpenAI](#)). On entend maintenant des jugements bien plus mesurés que les réactions alarmistes qui ont suivi la sortie d'AlphaGo : "C'est assez bluffant, mais il ne faut pas oublier qu'il s'agit d'un réseau de neurones artificiel, entraîné pour accomplir une tâche ; il n'y a aucune créativité ni aucune forme d'intelligence." (Fabienne Chauvière, [France Inter](#), 31 janvier 2021)

Aucune forme d'intelligence ? Ne soyons pas trop exigeants, mais restons lucides sur l'énorme fossé qui sépare les réseaux de neurones de ce que serait une véritable intelligence artificielle.

Jean-Louis Dessalles a écrit "Des intelligences très artificielles" aux éditions Odile Jacob (2019). Il est maître de Conférences, Institut Mines-Télécom (IMT)